Multimarginal optimal transport, density functional theory, and convex relaxation

Lexing Ying
Stanford University and FAIR

Joint work with Yuehaw Khoo

Jan 2019, BIRS

Optimal transport

- Given probability distributions ρ_1, ρ_2 on X
- Cost function $c(x, y)$
- Optimal transport problem

$$\inf_{\mu \in \Pi(\rho_1, \rho_2)} \int_X \int_X c(x, y) d\mu(x, y)$$

$\Pi(\rho_1, \rho_2)$ is the set of distributions on $X \times X$ with marginals ρ_1, ρ_2.

- Many applications:
  - Definition of the so-called Wasserstein distance,
  - Operational research, ...
  - Generative adversarial network (GAN) ...

Multi-marginal optimal transport

- Given marginals $\rho_1, \ldots, \rho_N$ on X
- Multimarginal OT problem

$$\inf_{\mu \in \Pi(\rho_1, \ldots, \rho_N)} \int_{X \times \cdots \times X} c(x_1, \ldots, x_N) d\mu(x_1, \ldots, x_N)$$

  with $\Pi(\rho_1, \ldots, \rho_N)$ the set of distributions on $X \times \cdots \times X$ with marginals $\rho_1, \ldots, \rho_N$
- Applications
  - Operational research, ...
  - Density functional theory
- Numerics: LP with exponential size in N
- Our goal: break this complexity barrier.

Density functional theory

- Many-body Schrödinger equation: finding ground state

$$\inf_{\psi} \int \psi(x_1, \ldots, x_N)^* H \psi(x_1, \ldots, x_N) dx_1 \ldots dx_N$$

  - H: Hamiltonian operator

$$H\psi \equiv \left(\sum_{i=1}^{N} -\Delta_{x_i} + \sum_{i<j} \frac{1}{|x_i - x_j|} + V_{\text{ext}} \right) \psi$$

  - $\psi(x_1, \ldots, x_N)$, $\|\psi\|_{\mathcal{L}_2} = 1$, antisymmetric.
- High dimensional problem, hard to solve.
- Density functional theory: Can change to a variational problem

$$\inf_{\rho} F[\rho] + V_{\text{ext}}[\rho]$$

  $\rho(\cdot)$: 1-marginal of $|\psi(x_1, \ldots, x_N)|^2$,
- $F[\rho]$: unknown universal functional (Hohenberg-Kohn 64).

$F[\rho]$: Strictly-correlated electron limit

- Usual approach: replace $F[\rho]$ with KS functional (Kohn-Sham 65)
- Opposite regime: approximate $F[\rho]$ with strictly-correlated electron (SCE) functional (Seidl 99)

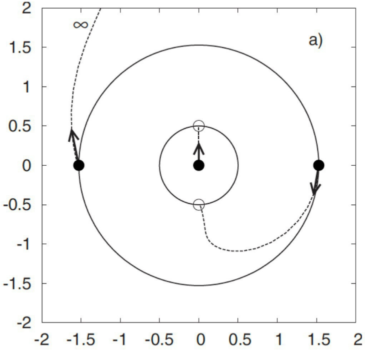
$$V_{\mathsf{ee}}^{\mathsf{SCE}}[\rho] = \inf_{\mu \in \Pi(\rho, \cdots, \rho)} \int \sum_{i<j} \frac{1}{|x_i - x_j|} \mu(x_1, \ldots, x_N) dx_1 \ldots dx_N$$

  with $\mu(x_1, \ldots, x_N) := |\psi(x_1, \ldots, x_N)|^2$ symmetric, $\int d\mu = 1, \mu \geq 0$, and $\mu(x_1, \ldots, x_N) = 0$ if any $x_i = x_j$.
- This is a special multimarginal OT problem
  - with same ρ for each dimension
  - Cost

$$c(x_1, \ldots, x_N) = \sum_{i<j} \frac{1}{|x_i - x_j|}.$$

SCE example

- Support of μ is singular.
- Li atom in SCE regime (Seidl-Gori-Giorgi-Savin 07).



Breaking the complexity barrier

- Numerics-related previous work
  - (Mendl-Lin 12): Solve the dual problem of $V_{ee}^{\mathsf{SCE}}[\rho]$ (exponential number of constraints)
  - (Benamou-Carlier-Nenna 16): Sinkhorn scaling (exponential number of variables)
  - (Friesecke-Vogler 18): Existence of sparse solution for multimarginal OT (optimization scheme to be worked out)

- Can we solve such a multimarginal OT with polynomial complexity?

- Approach: use convex relaxation techniques to obtain useful lower and upper bounds for the SCE optimization problem

Outline

- A lower bound to $V_{ee}^{\mathsf{SCE}}[\rho]$
- An upper bound to $V_{ee}^{\mathsf{SCE}}[\rho]$

Discretization

- For SCE: $c(x_1, \ldots, x_N) = \sum_{i<j} \frac{1}{|x_i - x_j|}$.
- Hence, focus on multimarginal OT with pairwise cost:

$$\inf_{\mu \in \Pi(\rho, \cdots, \rho)} \int_{X \times \ldots \times X} \sum_{k,l=1,k<l}^{N} c(x_k, x_l) d\mu(x_1, \ldots, x_N)$$

- Discretize X with L grid points $p_1, \ldots, p_L$ and redefine

$$c(i, i') \equiv \frac{1}{|p_i - p_{i'}|}.$$

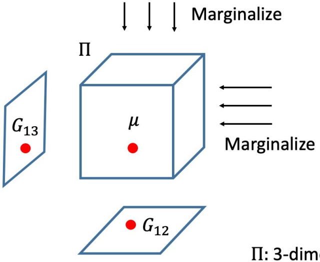
- From now on, focus on the discrete problem

$$\min_{\mu \in \Pi(\rho, \cdots, \rho)} \sum_{i_1, \ldots i_N = 1}^{L} \sum_{k,l=1,k<l}^{N} c(i_k, i_l) \mu(i_1, \ldots, i_N)$$

Reducing dimensionality

$$\min_{\mu \in \Pi(\rho, \cdots, \rho)} \sum_{i_1, \ldots i_N = 1}^{L} \sum_{k,l=1, k<l}^{N} c(i_k, i_l)\mu(i_1, \ldots, i_N)$$

- Rewrite the problem in terms of 2-marginals



Π: 3-dimensional probability polytope

Representing G_{kl} with μ

- Extreme points of density: delta functions.
- Write μ as convex combination of extreme points:

$$\mu = \sum_{i_1,\dots i_N} \mu_{i_1,\dots,i_N} e_{i_1} \otimes \cdots \otimes e_{i_N}$$

- $\{e_1, \dots, e_L\}$ are canonical basis vectors.
- $\sum_{i_1,\dots i_N} \mu_{i_1,\dots,i_N} = 1$ and $\mu_{i1,\dots,i_N} \geq 0$.
- The 2-marginal of the (k,l) slice: an $L \times L$ matrix

$$G_{kl} = \sum_{i_1,\dots i_N} \mu_{i_1,\dots,i_N} e_{i_k} e_{i_l}^{\mathsf{T}}$$

- Due to symmetry, all G_{kl} for $k \neq l$ are the same and all G_{kk} are the same

$$\gamma \equiv G_{kl}, \quad \epsilon \equiv G_{kk}$$

Equivalent multimarginal OT form

- Introduce $G \in \mathbb{R}^{NL \times NL}$:

$$G := \begin{bmatrix} G_{11} & \cdots & G_{1N} \\ \vdots & \ddots & \vdots \\ G_{N1} & \cdots & G_{NN} \end{bmatrix} = \begin{bmatrix} \epsilon & \cdots & \gamma \\ \vdots & \ddots & \vdots \\ \gamma & \cdots & \epsilon \end{bmatrix}$$

- Recall

$$\min_{\mu \in \Pi(\rho, \cdots, \rho)} \sum_{k,l=1, k<l}^{N} \sum_{i_k, i_l} c(i_k, i_l) \sum_{\text{other } i} \mu(i_1, \ldots, i_N)$$

- Write the optimization problem in terms of 2-marginals:

$$\min_{G \sim \Pi(\rho, \ldots, \rho)} \mathrm{Tr}(CG), \quad \text{with} \quad C = \begin{bmatrix} 0 & c & \cdots & c \\ c & 0 & & \\ \vdots & & \ddots & \vdots \\ c & & \cdots & 0 \end{bmatrix}$$

- Discrete quadratic optimization problem. Relax the domain of G.

Convex relaxation

- Problem: $\min_{G\sim\Pi(\rho,\ldots,\rho)} \mathrm{Tr}(CG)$
- Convexly relax $G := \begin{bmatrix} G_{11} & \cdots & G_{1N} \\ \vdots & \ddots & \vdots \\ G_{N1} & \cdots & G_{NN} \end{bmatrix} = \begin{bmatrix} \epsilon & \cdots & \gamma \\ \vdots & \ddots & \vdots \\ \gamma & \cdots & \epsilon \end{bmatrix}$
- Some necessary conditions
  - $G_{ij}\mathbf{1} \equiv \gamma\mathbf{1} = \rho$
  - $G_{ii} \equiv \epsilon = \mathrm{diag}(\rho)$,
  - $G \geq 0$, $G \succeq 0$,
- Drop all other constraints and obtain the convex problem

$$\min_{G=[G_{ij}]} \mathrm{Tr}(CG), \quad C = \begin{bmatrix} 0 & c & \cdots & c \\ c & 0 & & \\ \vdots & & \ddots & \vdots \\ c & & \cdots & 0 \end{bmatrix}$$

s.t. $G_{ij}\mathbf{1} \equiv \gamma\mathbf{1} = \rho,\ G_{ii} \equiv \epsilon = \mathrm{diag}(\rho),\ G \geq 0,\ G \succeq 0.$

Final SDP form

- Rewrite the cost

$$\mathrm{Tr}(CG) = \frac{N(N-1)}{2}\,\mathrm{Tr}(c\gamma)$$

- Finally, introduce a mix of 2 marginals and 1 marginal

$$\delta = \frac{1}{N^2}\begin{bmatrix} I \cdots I \end{bmatrix} G \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix} = \frac{1}{N^2}\begin{bmatrix} I \cdots I \end{bmatrix} \begin{bmatrix} \epsilon & \cdots & \gamma \\ \vdots & \ddots & \vdots \\ \gamma & \cdots & \epsilon \end{bmatrix} \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}$$

- Then

$$\delta = \frac{1}{N}\epsilon + \frac{N-1}{N}\gamma, \quad \delta\mathbf{1} = \rho, \quad \gamma = \gamma(\delta) = \frac{1}{N-1}(N\delta - \mathsf{diag}(\delta\mathbf{1}))$$

- A convex-relaxed SDP lower bound:

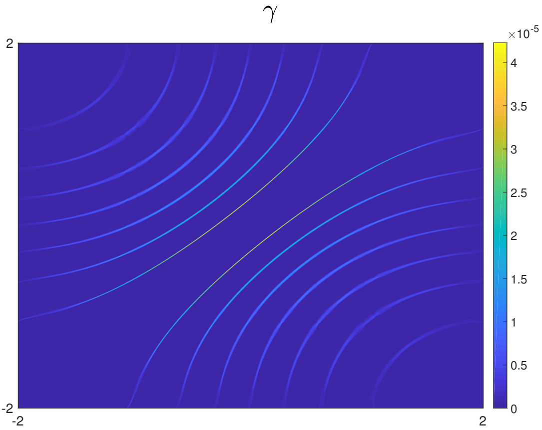
$$V_{ee}^{\mathsf{SCE}}[\rho] \approx \min_{\delta:\delta\mathbf{1}=\rho} \frac{N(N-1)}{2}\,\mathrm{Tr}\left(c\left(\frac{N}{N-1}\delta - \frac{1}{N-1}\mathsf{diag}(\delta\mathbf{1})\right)\right)$$

with $\delta \succeq 0, \quad \delta \geq 0, \quad \mathsf{diag}(\delta) = \frac{\delta\mathbf{1}}{N} = \frac{\rho}{N}.$

Example

- 1D electron: $N = 8$, $L = 1600$

$$\rho \propto \exp\left(-x^2/\sqrt{\pi}\right).$$



- 10^{25} entries if LP was used.

Why this relaxation is reasonable

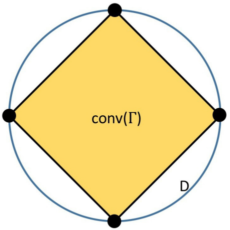
- Theorem (Friesecke-Vogler 18): The set of extreme points of N-representable symmetric 2-marginals (with Coulombic cost) is

$$\Gamma = \left\{ \frac{N}{N-1}\lambda\lambda^{\mathsf{T}} - \frac{1}{N-1}\mathrm{diag}(\lambda) \;\middle|\; \lambda \in \left\{0, \frac{1}{N}\right\}^{L},\ \lambda^{\mathsf{T}}\mathbf{1} = 1 \right\}$$

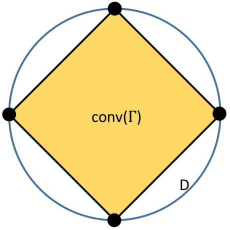
- Theorem (Khoo-Y. 18): Γ is a subset of the extreme points of

$$D \equiv \left\{ \frac{N}{N-1}\delta - \frac{1}{N-1}\mathrm{diag}(\delta\mathbf{1}) \;\middle|\; \delta \succeq 0,\ \delta \geq 0,\ \mathrm{diag}(\delta) = \frac{\delta\mathbf{1}}{N} \right\}$$

(Note that D is the set of feasible 2-marginals γ for the SDP).



Why this relaxation is reasonable



- Thus $\gamma \in \text{conv}(\Gamma) \subset D$
- If the relaxation is sufficiently tight, we expect for δ^* (the minimizer of the relaxed problem) to satisfy:

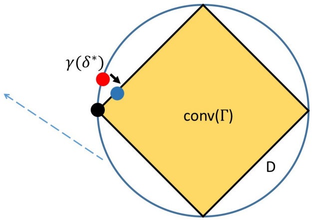
$$\delta^* \approx \sum_{g=1}^{m} a_g \lambda_g \lambda_g^\mathsf{T}, \quad \lambda_g \in \left\{0, \frac{1}{N}\right\}^L, \quad \lambda_g^\mathsf{T}\mathbf{1} = 1.$$

Outline

- A lower bound to $V_{ee}^{\mathsf{SCE}}[\rho]$
- An upper bound to $V_{ee}^{\mathsf{SCE}}[\rho]$

Upper bound

- With constraint $\delta \mathbf{1} = \rho$, solution $\gamma^* \equiv \gamma(\delta^*)$ usually out of $\text{conv}(\Gamma)$.



- Need to project $\gamma^* \equiv \gamma(\delta^*)$ back, or equivalently write

$$\delta^* \to \sum_{g=1}^{m} a_g \lambda_g \lambda_g^{\mathsf{T}},$$

with $\lambda_g \in \{0, \frac{1}{N}\}^L$ for $g = 1, \ldots, m$, (here $m = L$).

- How to perform the projection?
  - One possibility: use eigendecomposition plus thresholding
  - Fails because $\{\lambda_g\}$ are non-orthogonal
  - Idea: use 3-marginals

Tensor decomposition with 3-marginals

- Recall that one needs to retrieve each λ_i.
- Solution: use 3-marginals (why?) and apply similar derivation

---

- Marginalize to 3-marginals
- Symmetrize the 3-marginals by averaging
- Use the N-representable 3-marginal θ (instead of 2-marginals δ) as optimization variable
- Apply a similar convex relaxation to $\theta \in \mathbb{R}^{L\times L\times L}$ as before
- Solve the multimarginal OT in terms of 3-marginals

- More expensive, but still independent of N and no exponential blowup

Why 3-marginal?

- CP-decomposition for θ^* (the minimizer of the relaxed problem)

$$\theta^* \Rightarrow \sum_{g=1}^{m} a_g \lambda_g \otimes \lambda_g \otimes \lambda_g$$

  and the RHS serves as the upper bound.
- For 3-tensor, one has unique decomposition results for $\lambda_g, g = 1, \ldots, m$ up to scaling under very mild condition.
- $m = L$ in our case.
  - Reason: there are $L - 1$ effective constraints $\delta \mathbf{1} = \rho$.
  - Each extra constraint increases the support by one.
  - So needs a convex combination of $1 + (L - 1) = L$ extreme points.

Projection of 3-marginals

- Apply Jenrich's algorithm (weighted sum in third dim). Choose random vectors w_1 and w_2

$$W_1 = \sum_{g=1}^{m} \theta^*(:,:,g) w_1(g), \quad W_2 = \sum_{g=1}^{m} \theta^*(:,:,g) w_2(g)$$

  Plug in $\theta^* = \sum_{g=1}^{m} a_g \lambda_g \otimes \lambda_g \otimes \lambda_g$

$$W_1 = \sum_{g=1}^{m} (a_g w_1^\top \lambda_g) \lambda_g \lambda_g^\top, \quad W_2 = \sum_{g=1}^{m} (a_g w_2^\top \lambda_g) \lambda_g \lambda_g^\top.$$

- $\{\lambda_g\}_{g=1}^m$ are linearly independent. By using $U = [\lambda_1 \cdots \lambda_m]$,

$$W_1 = U \Sigma_1 U^\top, \quad W_2 = U \Sigma_2 U^\top.$$

$$W_1 W_2^{-1} = U \Sigma U^{-1}, \quad \Sigma = \text{diag}\left(\left[\frac{a_1 w_1^\top \lambda_1}{a_1 w_2^\top \lambda_1}, \ldots, \frac{a_m w_1^\top \lambda_m}{a_m w_2^\top \lambda_m} \right] \right).$$

- Eigendecomposition of $W_1 W_2^{-1}$ gives U and $\{\lambda_g\}_{i=1}^m$

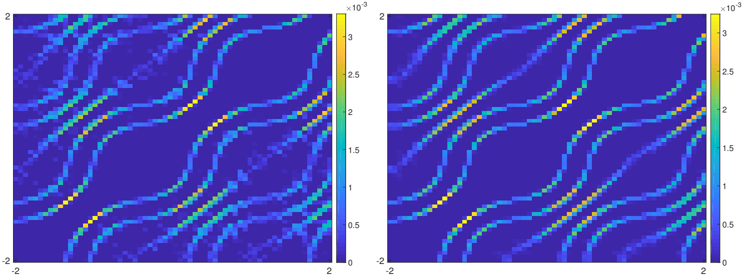
Numerical examples

- Lower bound: obtained from 2-marginal δ^* or 3-marginal θ^*.
- Upper bound: always obtained from 3-marginal θ^*.

Numerical examples

- 1D electrons, $N = 8$.

$$\rho(x) \propto \sin(4x) + 1.5.$$

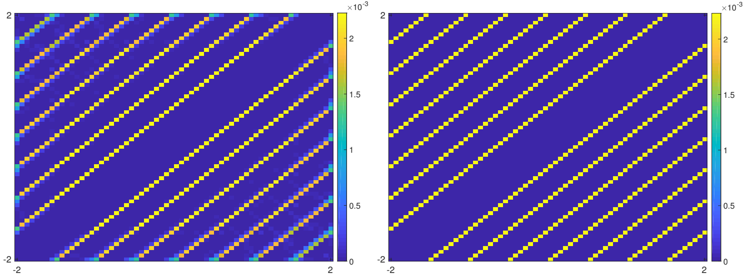


- Left: 2-marginal $\delta^* \to \gamma$. Relative gap = 4.2e-02
- Right: 3-marginal $\theta^* \to \gamma$. Relative gap = 3.9e-02

Numerical examples

- 1D electrons, $N = 8$.

$$\rho(x) \propto 1.$$

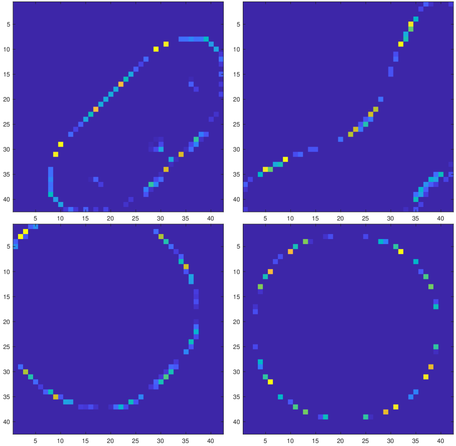


- Left: 2-marginal $\delta^* \to \gamma$. Relative gap $=$ 4.9e-04
- Right: 3-marginal $\theta^* \to \gamma$. Relative gap $=$ 1.0e-06

Numerical examples

- 2D electrons, $N = 5$.

$$\rho(x, y) \propto 1.$$



- Plots are slice of 2-marginal with one component fixed.
- 2-marginal $\delta^* \to \gamma$. Relative gap = 3.8e-02
- 3-marginal $\theta^* \to \gamma$. Relative gap = 3.5e-02

Thank you

- Email: lexing@stanford.edu
- URL: http://web.stanford.edu/~lexing
- Research supported by NSF and DOE